

PATENT APPLICATION
Attorney Docket No. 8371-148

INTEGER COSINE TRANSFORM MATRIX FOR PICTURE CODING

BACKGROUND

This application claims priority from U.S. Provisional Application Serial No. 60/255,352, filed December 13, 2000.

Integer-based transform matrices are used for transform coding of digital signals, such as for coding image/video signals. Discrete Cosine Transforms (DCTs) are widely used in block-based transform coding of image/video signals, and have been adopted in many Joint Photographic Experts Group (JPEG), Motion Picture Experts Group (MPEG), and network protocol standards, such as MPEG-1, MPEG-2, H.261, and H.263. Ideally, a DCT is a normalized orthogonal transform that uses real-value numbers. This ideal DCT is referred to as a real DCT. Conventional DCT implementations use floating-point arithmetic that require high computational resources. To reduce the computational burden, DCT algorithms have been developed that use fix-point or large integer arithmetic to approximate the floating-point DCT. However, none of these approaches has been able to guarantee coding reversibility. Coding reversibility refers to the ability of a transform algorithm to transform a signal and then inverse transform the transformed signal as closely as possible back to the original signal, without inducing error into the original signal.

Integer transform techniques have been developed to provide coding reversibility. Some of these techniques are described in the following documents. US patent 5,999,957, Ohta, (2000) “Lossless Transform Coding System For Digital Signals,” assigned to NEC Corporation which is a division of US patent 5,703,799. US patent 5,999,656, Zandi and Schwartz, (1999) “Overlapped Reversible Transforms for Unified Lossless/Lossy Compression,” assigned to Ricoh Corporation. US patent 5,703,799, Ohta, (1997) “Lossless Transform Coding System For Digital Signals,” assigned to NEC Corporation. Ying-Jui Chen, Soontorn Oraintara, and Truong Nguyen, “Video Compression Using Integer DCT,” Proceedings of IEEE International Conference on Image Processing (ICIP), 2000.

Among these techniques, transform matrices are derived based on a Hadamard transform to approximate the real DCT. Usually, small integer transform coefficients are chosen to improve coding efficiency. However, the Hadamard transform is a redundant transform in the frequency (or real DCT) domain. These transform matrices were developed without considering the distortion in the frequency domain with respect to the real DCT. The normalization of the transform matrix is also not considered.

SUMMARY OF THE INVENTION

An integer transform matrix is used for implementing a Discrete Cosine Transform (DCT). Optimized values for the integer transform matrix are derived that satisfy certain normalization constraints and that also minimize the frequency distortion in the transform matrix.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a representation of a 4 x 4 integer matrix.

Fig. 2 shows a representation of a 8 x 8 integer matrix.

Fig. 3 shows a representation of a 16 x 16 integer matrix.

Fig. 4 shows a block diagram that shows an image processing system that uses the integer matrices shown in Figs 1-3.

DETAILED DESCRIPTION

Different sized blocks in an image can be coded with a discrete cosine transform. For example, 16 x 16, 16 x 8, 8 x 16, 8 x 8, 8 x 4, 4 x 8, and 4 x 4 macroblocks can be transformed using horizontal and vertical transform matrices of size 4 x 4, 8 x 8, and 16 x 16. The macroblocks are transformed according to the following:

$$C_{n \times m} = T_m \times B_{n \times m} \times T_n^T,$$

where $B_{n \times m}$ denotes an image block with n pixels and m rows. The terms T_n and T_m represent the horizontal and vertical transform matrices of size $n \times n$ and $m \times m$, respectively. The term $C_{n \times m}$ denotes the cosine transformed $n \times m$ block.

An orthogonal 4×4 integer matrix $T_m = T_4$ can be written based on a Hadamard transform as shown in FIG. 1. The values n_0, n_1, n_2 , and n_3 represent integers. Matrix normalization requires the constraints shown in equation 1, where *norm* is an integer representing the normalization factor of the matrix.

$$\begin{cases} n_0 = \text{norm} \\ n_1^2 + n_3^2 = 2 \cdot \text{norm}^2 \\ n_2 = \text{norm} \end{cases}$$

EQUATION (1)

The real DCT of the base vector t_i is represented as d_i as shown in equation 2.

$$d_i = t_i \cdot DCT$$

EQUATION (2)

DCT represents the real DCT matrix, which contains the basis functions of the real DCT in column vectors. The frequency distortion of the transform matrix with respect to the real DCT is then defined in equation 3.

$$E_4 = \frac{1}{4} \sum_{i=0}^4 \sum_{\substack{j=0 \\ j \neq i}}^4 \frac{|d_i(j)|}{|d_i(i)|}$$

EQUATION (3)

For each factor *norm*, the optimal transform matrix, if any, is the one giving the minimum distortion as defined in equation (3). Since a large value is expected for $|d_i(i)|$, a matrix with any zero value $d_i(i)$ will be ruled out. This can be taken as an implicit requirement by the definition.

The integer coefficients of several optimal 4×4 transform matrices satisfying equation (1) and having the minimum frequency distortion as defined in equation (3) are listed in the following table. The corresponding matrix with *norm* of 13 is being used in International Telecommunication Union Standard ITU-T H.26L. The information shown in the following table reveals that a *norm* of 13 is the most reasonable choice.

n_0, n_1, n_2, n_3	<i>norm</i>	DCT distortion
13,17,13,7	13	0.11%
17,23,17,7	17	4.88%
25,17,25,31	25	5.47%

An 8×8 transform matrix is shown in Fig. 2. The normalization constraints in this case are shown in equation 4.

$$\left\{ \begin{array}{l} n_0 = \text{norm} \\ n_1^2 + n_3^2 + n_5^2 + n_7^2 = 4 \cdot \text{norm}^2 \\ n_2^2 + n_6^2 = 2 \cdot \text{norm}^2 \\ n_4 = \text{norm} \end{array} \right.$$

EQUATION (4)

The corresponding distortion function is shown in equation 5.

$$E_8 = \frac{1}{8} \sum_{i=0}^8 \sum_{\substack{j=0 \\ j \neq i}}^8 \frac{|d_i(j)|}{|d_i(i)|}$$

EQUATION (5)

The integer coefficients of an optimal 8×8 orthogonal transform matrix satisfying the normalization constraints in equation 4 and having the minimum frequency distortion for equation 5 are listed in the following table.

n_0, n_1, \dots, n_7	<i>norm</i>	DCT distortion
17,24,23,20,17,12,7,6	17	6.95%

The matrix with *norm* of 17 is currently being used in the H.26L ABT Core Experiment. A 16×16 matrix is shown in Fig. 3. The normalization constraints for the 16×16 matrix are shown in equation 6.

$$\left\{ \begin{array}{l} n_0 = \text{norm} \\ n_1^2 + n_3^2 + n_5^2 + n_7^2 + n_9^2 + n_{11}^2 + n_{13}^2 + n_{15}^2 = 8 \cdot \text{norm}^2 \\ n_2^2 + n_6^2 + n_{10}^2 + n_{14}^2 = 4 \cdot \text{norm}^2 \\ n_4^2 + n_{12}^2 = 2 \cdot \text{norm}^2 \\ n_8 = \text{norm} \end{array} \right.$$

EQUATION (6)

The corresponding distortion function for the 16×16 matrix is shown in equation 7.

$$E_{16} = \frac{1}{16} \sum_{i=0}^{16} \sum_{\substack{j=0 \\ j \neq i}}^{16} \frac{|d_i(j)|}{|d_i(i)|}$$

EQUATION (7)

The integer coefficients for an optimal 16×16 orthogonal transform matrix were derived using the constraints in equation 6 and minimizing the distortion function in equation 7. The optimal matrix values derived for the integer coefficients are shown in the following table.

n_0, n_1, \dots, n_{15}	<i>Norm</i>	DCT distortion
17, 22, 24, 28, 23, 12, 20, 20, 17, 12, 12, 16, 7, 8, 6, 6	17	32.77%

In summary, to derive an optimal $2^m \times 2^m$ transform matrix, there should be $(m+1)$ normalization constraints.

$$\left\{ \begin{array}{l} n_0 = \text{norm} \\ \sum_{i=0}^{2^{m-1}-1} n_{2i+1}^2 = 2^{m-1} \cdot \text{norm}^2 \\ \sum_{i=0}^{2^{m-2}-1} n_{4i+2}^2 = 2^{m-2} \cdot \text{norm}^2 \\ \sum_{i=0}^{2^{m-3}-1} n_{8i+4}^2 = 2^{m-3} \cdot \text{norm}^2 \\ \vdots \\ n_{2^{m-1}} = \text{norm} \end{array} \right.$$

EQUATION (8)

And for a specific *norm*, the optimal orthogonal matrix should minimize the DCT distortion function shown in equation 9.

$$E_{2^m} = \frac{1}{2^m} \sum_{i=0}^{(2^m-1)} \sum_{\substack{j=0 \\ j \neq i}}^{(2^m-1)} \frac{|d_i(j)|}{|d_i(i)|}$$

EQUATION (9)

Equation 9 effectively multiplies the real DCT with the integer cosine transform matrix. The ideal result is a normalized matrix with zero values everywhere except along the diagonal of the resultant matrix. The diagonal coefficients being 1's. Equation 9 identifies the total error generated by the non-zero values in the nondiagonal coefficients in the resultant between the real DCT and the integer cosine transform.

Since the transform matrix derived in this manner has minimum distortion with respect to a real DCT, it can be used with many existing DCT-based coding techniques. For example, the transform matrix can be used with frequency-based or HVS-based video coding, such as scanning, quantization, and filtering. There is flexibility in defining the distortion function. For

example, weighting factors can be assigned to each frequency component based on an HVS mapping.

FIG. 4 shows one example of a system 10 that encodes and decodes data using the optimized transform matrices shown above. The system 10 can be any computer, video device, camera, network processing device, etc. that processes data. Data in block 12 can be any type of information that needs to be transformed. In one example, the system 10 processes video data.

A Discrete Cosine Transform (DCT) in block 14 uses one of more of the optimized transform matrices shown above to DCT transform the data from block 12. The size of integer matrices used on the data depends on size of the blocks the image data is sectioned into. For example, in one application, image data may be sectioned into 4 x 4 bit macroblocks. In the same or another application, it may be determined that another section of the same image, or a different image, can be more efficiently coded by using 16 x 16 bit macroblocks. The DCT block 14 includes a memory that contains the different 4 x 4, 8 x 8, and 16 x 16 optimized transform matrices described above. The transform matrix corresponding to the image block size is applied to the individual macro blocks in the image data.

The transformed data is quantized in a block 16 and then variable length coded (VLC) in block 18. The encoded data is either stored in a memory or transmitted over a communication channel in block 20.

The data is decoded by first inverse variable length coding the data in block 26 and inverse quantizing (IQ) the data in block 25. An Inverse Discrete Cosine Transform (IDCT) in block 24 uses the optimized integer cosine matrices for inverse cosine transforming the decoded data. The inverse cosine transform is implemented by applying inverse integer matrices for the matrices shown in Figs. 1-3. For example, the inverse cosine transform is generated according to the following:

$$B_{n \times m} = T_m^T \times C_{n \times m} \times T_n,$$

where $B_{n \times m}$ denotes the inverse transformed image block with n pixels and m rows, T_n and T_m represent the horizontal and vertical integer transform matrices of size $n \times n$ and $m \times m$, respectively, and $C_{n \times m}$ denotes the cosine transformed $n \times m$ image block.

The inverse integer matrices are selected to correspond to the block sizes used to originally transform the data. The inversed transformed data is then output as inverse transformed data in block 20.

The system described above can use dedicated processor systems, micro controllers, programmable logic devices, or microprocessors that perform some or all of the operations. Some of the operations described above may be implemented in software and other operations may be implemented in hardware.

For the sake of convenience, the operations are described as various interconnected functional blocks or distinct software modules. This is not necessary, however, and there may be cases where these functional blocks or modules are equivalently aggregated into a single logic device, program or operation with unclear boundaries. In any event, the functional blocks and software modules or described features can be implemented by themselves, or in combination with other operations in either hardware or software. Having described and illustrated the principles of the invention in a preferred embodiment thereof, it should be apparent that the invention may be modified in arrangement and detail without departing from such principles. Claim is made to all modifications and variation coming within the spirit and scope of the following claims.